

Agilent G2721AA/G2733AA Spectrum Mill MS Proteomics Workbench

Quick Start Guide

What is the Spectrum Mill MS Proteomics Workbench?	2
What's New in Version B.06.00?	3
Where to Find More Information	9
Setting Up Spectrum Mill Server and Client	11
Transferring Spectral Files to Spectrum Mill Server	14
Starting the Spectrum Mill Software	19
Using Normal and Iterative Workflows to Process Data	21
Running Workflow Automation or Interactive Automation	25
Quick Reference	28
Spectrum Mill modules and their purposes	28
Interpreting scores from MS/MS Search	31
Protein/Peptide Summary Modes	33

Use this guide for your first steps with the Spectrum Mill workbench.



Agilent Technologies

What is the Spectrum Mill MS Proteomics Workbench?

The Spectrum Mill software is a collection of tools for high-throughput processing of MS/MS and MS spectra to provide protein and peptide identifications and relative quantitation. The Spectrum Mill workbench can do in minutes what could take you hours or days to do manually.

This software includes bioinformatics tools for:

- Spectral preprocessing for MS/MS and MS-only data, including spectral quality filtering and precursor charge state assignment for MS/MS spectra
- Protein database search, including searches for both MS/MS and MS-only spectra
- Data review, validation, and comparison for large data sets, including those from two-dimensional LC/MS/MS analyses and differential expression studies
- *De novo* sequencing
- Identification of mutations, post-translational modifications, and chemical modifications
- Differential expression quantitation

The Spectrum Mill workbench accepts Agilent data, as well as data from certain other vendors' mass spectrometers. The software executes on a web server, and you access the program via a web browser window.

The software also includes process automation tools to automate the preprocessing, search, validation and review steps.

What's New in Version B.06.00?

The latest B.06.00 version of Agilent Spectrum Mill MS Proteomics Workbench includes performance improvements and other enhancements that increase the flexibility of the software and enable greater productivity. This section describes differences between the B.05.00 and the B.06.00 versions of the Spectrum Mill workbench. See the online Help for features added in previous revisions.

Installation and configuration

The following installation and configuration features are new in the Spectrum Mill workbench version B.06.00. For details, see the *Installation Guide* and the *Site Preparation Guide*. The latter is on your software disk.

Operating system

Spectrum Mill B.06.00 is supported on the following 64-bit operating systems:

- Windows 7 and Windows Server 2008 R2
- Windows Server 2012 R2
- Windows 10
- Windows Server 2016

Only 64-bit versions of the above operating systems are supported.

Windows XP and Windows Server 2003 are no longer supported.

If you configure a new system, Agilent recommends that you provide adequate memory (16 GB or more) and adequate disk space (1–2 TB for data). For a new system, Agilent recommends 24 to 32 GB of memory.

Perl

ActiveState Perl 5.18.4 is supplied with this release. (If you are updating from Spectrum Mill B.04.01 or earlier, uninstall the prior Perl version first.)

What's New in Version B.06.00?

Installation and configuration

Java (JRE)

Java JRE 1.8u111 is included in this release. Install both the 32- and 64-bit JREs, even if you run the 64-bit Internet Explorer (IE 11).

JRE 8.u111 is required to support the Spectrum Viewer and other applets.

Be sure to install JRE on all client browsers, as well as the Spectrum Mill server.

Java updates are regularly released by Oracle to address security issues. In general, you should be able to install these updates without issue. Use the MS Isotope utility to test if the version is compatible with the Spectrum Viewer applets. The spectra are not displayed if there is a problem with the applets.

If the Spectrum Viewer shows no content, or does not display properly, see the Help for Tips and Tricks. Specifically, see the section on “Troubleshooting - Spectrum Viewer.”

IIS

IIS must be installed before you install the Spectrum Mill workbench.

See the *Installation Guide* for details on installation and configuration of IIS.

MSFileReader support (for .raw files)

To process ThermoFisher Scientific .raw files, Spectrum Mill B.06.00 can use the 64-bit ThermoFisher Scientific MSFileReader. To use this program, install it on the server before or after you install the Spectrum Mill workbench.

If you update your Spectrum Mill installation, and you previously had installed the 32-bit MSFileReader, it must be updated to the 64-bit version.

Web browser support

The following Web browser is supported with this version:

- Microsoft Internet Explorer 11

If you are upgrading from Spectrum Mill workbench B.04.01 or earlier, note that the prior release required Compatibility View for Internet Explorer 9 and later. For this release you must disable Compatibility View.

MS/MS Search

- MS/MS Search can be from two to five times faster, depending upon the peptide redundancy in the databases you search.
- MS/MS Search no longer creates individual **spo** files in **results_mstag**. Instead, searches create temporary concatenated **spo** files (**cpo** files). The **spo** files are added to an **spo.zip** file. The program deletes the individual **cpo** files when a search completes.
 - This improvement significantly reduces the load on the file system (both in space and number of files), and makes archiving and copying data folders much faster.
- In the Spectral Quality Filtering section of the MS/MS Search page, the **min# of peaks** filter has been removed and **Precursor isotope quality** and **Precursor isolation purity** filters have been added.
- HLA peptide motifs and related half-enzyme searches are now supported.
- The C-terminal peptide can now be matched in half-enzyme digests that are built C-term to N-term. Previously, the C-terminal peptide would only be matched, if it matched the enzyme specificity.
- By default, peptides with lengths less than five amino acids are no longer matched. This enhancement is especially valuable when you allow large precursor mass shifts for variable modifications. To enable matching of smaller peptides, mark the **Dynamic peak thresholding** check box.
- If you search data with multiple fragmentation modes and you select **All**, an informative error message is generated to indicate which mode(s) should be selected for searching.

Quality Metrics and False Discovery Rate (FDR)

- Additional metrics are now available.
- The defect fix that was in B.05.00 SP1 is included. In Spectrum Mill B.05.00, the reversed hit with the second best score was sometimes reported instead of the one with the best score. This caused the search to underestimate FDR. If you search against a typical UniProt database, the defect fix reduces the total number of identifications by 1-5%, but it makes them more accurate.

Home page

- Spectrum Matcher is back (under Mass Spectral Interpretation Tools) and provides additional quality filtering features to assist in evaluating instrument performance or method changes.

Protein Databases

- Addition of tool to **Concatenate FASTA files**, which allows you to link together two or more existing databases. It includes two possible ways to concatenate:
 - Select one or more existing database files to concatenate.
 - Concatenate all files in a folder.
- Support for change of NCBI FASTA header format:
 - In September 2016, NCBI changed the FASTA header format to only supply the **gb** (GeneBank) accession. The former **gi** accession is no longer indicated.
 - Newly downloaded databases in the new format are supported and the **gb** accession is used by Spectrum Mill for those databases.
 - For the Spectrum Mill workbench to properly recognize the format, these new databases require either an **NCBIgb** or **gb** prefix instead of the **NCBInr** prefix.
 - Existing databases (**NCBInr**) are still supported. GeneBank accessions (when present) can be reported in Protein/Peptide Summary by creating a Category file for the database.

Tool Belt

- Now supports TMT 10-plex reporter ion correction factors
- When you remove prior search results, any applied correction factors remain.

Spectrum Matcher

- Includes new spectral quality filters and homology/variable mode
- Supports Load/Save parameter files (but not in workflows)

Protein/Peptide Summary

- Enhancements to support ion mobility data:
 - Includes new Ion Mobility review field to report DT (drift time in milliseconds) and CCS (collision cross section in square angstroms)
 - Agilent Proteomics Results (APR) Export can now export DT and CCS values if they are present in ion mobility data.
- Review Fields now include Correction Method (for Reporter Ratios). Select **Apply** to apply correction factors or **None** to skip correction.

Spectrum Summary

- Includes spectral quality filters
- Supports Load/Save parameter files
- New Review Field for ion mobility: reports DT and CCS values if available

Data extractors

- Thermo .RAW Extractor now supports Thermo Fusion/Lumos data.
- Generic Extractor now extracts to mzXML.
- Generic Extractor now supports PKL files from IM-MS Browser. (IM is ion mobility.)
 - The IM-MS Browser Rev 7.02 Build 209 or later can export concatenated PKL files that contain the retention time (RT), drift time (DT), and collision cross section (CCS) for a feature (precursor).
 - To report the CCS value, the CCS Single Field calibration must be applied to the data.
 - The names of the extracted spectra indicate RT (in minutes x 10) and DT (in msec x 10), so the name has the form myDataFile.<RTx10>.<DTx10>.<charge>.pkl.
 - This name provides a way to more easily find the spectrum in IM-MS Browser.
 - The RT, DT, and CCS values are stored in the specFeatures.tsv file, and reported in Protein/Peptide Summary when the **Ion mobility** review field is marked.
 - DT and CCS values from ion mobility experiments are written to mzXML for import into Skyline.

Where to Find More Information

Scientists: Read these guides and online help

Familiarization Guide Follow step-by-step instructions to process example data.

Application Guide Learn details to use the software.

Online Help Consult the online help for in-depth information not given in the *Application Guide*. To access help, click links on the home page, **Help** buttons on Spectrum Mill pages, or blue dividing bars on Spectrum Mill pages.

System Administrators: Read these guides and online help

Site Preparation Guide Use this guide to prepare for the software installation.

Installation Guide Use this guide to install the software on the server.

Application Guide See the following chapters:

- Chapter 9: System Administration
Get an overview on how to install databases and do other system administration tasks.
- Chapter 10: Files Created during Spectrum Mill Data Processing
Refer to this chapter to troubleshoot data processing, to selectively remove parts of the processing, or to decide which files to archive.

Online Help From any online Help page, click links under **For System Administrators**:

- Protein Databases (link to millhtml\SM_instruct\faman.htm)
Learn details on how to install databases, create indices, and create subset databases.
- Server Administration (link to millhtml\SM_instruct\servadmn.htm)
Learn details to do other system administration tasks.

Contact Agilent Technical Support

Internet, phone and e-mail

Internet: In the upper right corner of the Spectrum Mill home page, click **Support**.

Phone: For Agilent's worldwide sales and support center telephone numbers, go to <http://www.agilent.com/chem/contactus>.

From the US and Canada, call 800-227-9770 (select options 3-2-2).

E-mail: For e-mail assistance, go to <http://www.agilent.com/chem/contactus>. Under **e-mail Assistance**, click **Support Related Inquiries**. Fill out the form and submit.

Software Status Bulletins and software patches

A list of known problems for the Spectrum Mill workbench, along with possible solutions, is found on the Agilent Web site at http://www.chem.agilent.com/cs/Library/Support/Patches/SSBs/Spectrum_Mill.html.

Setting Up Spectrum Mill Server and Client

To set up the server

See the *Spectrum Mill MS Proteomics Workbench Installation Guide*. If you wish to update from a previous version of the Spectrum Mill workbench, see Chapter 2. Note that the server name cannot have an underscore.

To set up your client PC

Operating system

Check that you have a 64-bit version of Windows. Client PCs can use any version of Windows that supports Internet Explorer 11.

Browser

Verify that Internet Explorer 11 is installed.

- 1 Open Internet Explorer.

If you cannot see the Internet Explorer menus, right-click in the space next to the last tab and select **Menu bar**. The menus now appear in the upper left of Internet Explorer.

Or

Press the **Alt** key to temporarily see the menus.

- 2 Select **Help > About Internet Explorer**.

Notes about Compatibility View

In versions of Spectrum Mill prior to B.05.00, it was necessary to configure the browser to use Compatibility View. Now, the opposite is true. You must *not* enable Compatibility View for the Spectrum Mill server. If pages do not display properly, it is likely that Compatibility View is enabled.

If you get a warning to disable Compatibility View when you access the Spectrum Mill server, do the following:

Setting Up Spectrum Mill Server and Client

To set up your client PC

- 1 In Internet Explorer, click **Tools > Compatibility View Settings**. If the Spectrum Mill server is listed under **Websites you've added to Compatibility View**, remove it.
- 2 Clear the check box for **Display intranet sites in Compatibility View**.

Java (JRE)

Check that Java SE is installed. If not, the Spectrum Mill disk provides a version of Java SE (1.8u111) that can be installed. The latest version of Java SE can be downloaded from

<http://www.oracle.com/technetwork/java/javase/downloads/index.htm>.

- 1 Select **Tools > Internet Options...**
Click **Programs**, then click **Manage add-ons**.
- 2 Scroll down to the **Java (Oracle)** section to view the version.

When you update Java, first uninstall the prior versions, then install the 32-bit version, then the 64-bit version.

Both 32-bit (i586) and 64-bit (x64) versions of JRE are provided. For 64-bit operating systems, install both the 32-bit and the 64-bit versions to allow use of the proper JRE with either version of Internet Explorer. (On 64-bit Windows, both 32-bit and 64-bit versions of Internet Explorer are installed.)

Be sure to install the JRE on all client PCs.

Screen resolution

Spectrum Mill works best with the highest screen resolution your monitor supports.

Cookies

Enable cookies so you can save page settings.

- 1 For detailed instructions to enable cookies, search Internet Explorer help for **cookies**.
- 2 If the drop-down menus in the Spectrum Mill workbench appear empty, you may need to add the Spectrum Mill server as a trusted site. Search Internet Explorer help for **trusted site**. Also, make sure Compatibility View is *not* enabled.

Active scripting

Enable JavaScript (Active Scripting). This step is not required for Windows 10.

- 1 Select **Tools > Internet Options...**
- 2 Click the **Security** tab.
- 3 Select **Local intranet**.
- 4 Click **Custom level...**
- 5 In the **Security Settings** dialog, scroll down until you see the **Scripting** section.
- 6 Under **Active scripting**, click the **Enable** option.

Transferring Spectral Files to Spectrum Mill Server

To set up a useful directory structure

Transferring Spectral Files to Spectrum Mill Server

Before you open the Spectrum Mill home page, transfer your data files to the Spectrum Mill server. Most data files require no special preparation; you merely move or copy the files into the appropriate directory on the Spectrum Mill server, as described below. Some other types of data files must be preprocessed on the manufacturer's data system to generate generic peak list files. You then copy or move peak list files as described below.

NOTE

Spaces and parentheses are not allowed in filenames, so replace them with a dash or underscore. The following characters are also not permitted: |, ; % < > ? + !. Dots or periods (.) are allowed only for file suffixes (e.g., **myfile.d**).

To set up a useful directory structure

To make it easy to compare data sets, it is important that you set up the appropriate directory structure for your files on the Spectrum Mill server. See [Figure 1](#).

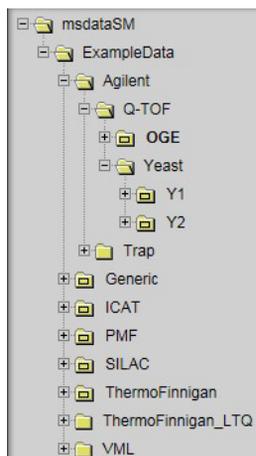


Figure 1 Default directory tree for Spectrum Mill data files

Whenever you want to compare samples in a set, you need to set up a subdirectory for each sample. This sample directory may contain data files from multiple sample fractions or gel slices. Here are some examples:

- If you run a two-dimensional LC/MS/MS analysis of a single sample, transfer all the files to a single directory on the server.
- If you run the same two-dimensional LC/MS/MS analysis on a second sample, or if you repeat the run on the first sample, transfer all these files to a second directory.
- For differential expression quantitation experiments where you use isotopic labels, transfer all the files to a single directory on the server. This is because all of the versions of the peptide (light, medium, heavy) are found in the same file or sets of files.
- If you conduct a differential expression study where you use the absolute abundances to do the quantitation, transfer samples from one cell state into one directory and the second cell state into a second directory.

To access the server directory msdataSM

When you transfer data files you move them to the **msdataSM** directory in the Data folder on the Spectrum Mill server.

Access the folder from the server

- In your Spectrum Mill file system on the server, navigate to the directory **msdataSM**.

This directory is often on the Data drive under the Spectrum Mill folder. (If you don't know how to find your Spectrum Mill file system, ask the person who installed your software.)

Access the folder from a PC client

- 1 Make sure your administrator has shared the **msdataSM** folder and given you permissions. See *Installation Guide*.
- 2 Map the shared folder, **\\servername\msdataSM**, to a client PC. (Right-click **Computer** and select **Map network drive**. Follow the instructions.)

To transfer MS/MS data files

To transfer raw data

- 1 In the directory **msdataSM**, create a new subdirectory for your sample files. Name it anything you want, but do not use spaces or parentheses in the name. (See “[To access the server directory msdataSM](#)” on page 15.) See the note on [page 14](#) about illegal characters in file names.

If you have multiple samples you wish to compare, remember to set up a subdirectory for each sample. Ensure that each subdirectory contains all fractions for a single sample.

You may add up to ten directory levels between **msdataSM** and your sample files, but shorter path lengths reduce memory usage and speed processing of large data sets.

- 2 Copy your *.d or *.raw files to your new subdirectory or subdirectories. (Note that *.d files are actually folders.)
- 3 After you transfer data files to the server, remove any spaces or parentheses in the data file names. Delete them or replace them with other characters, such as underscores.
- 4 Make sure you have both read and write permissions for each subdirectory that contains your sample files. (To check, right-click the folder and select **Properties**.) Make sure all user groups have full permissions. (Check the **Security** tab.)

To prepare and transfer generic peak list files (e.g., Waters Micromass Q-ToF or MALDI-TOF-TOF data or IM-MS Browser exported peak lists)

- 1 Prepare a generic peak list file.
 - For Waters Micromass Q-ToF, export centroided data with the Waters Micromass ProteinLynx package. The results are individual *.pkl or *.dta spectral files, or appended *.pkl files that contain multiple spectra.
 - For MALDI-TOF-TOF, export *.mgf files.
- 2 In the directory **msdataSM**, create a new subdirectory for your sample files. Name it anything you want, but do not use spaces or parentheses in the name. (See “[To access the server directory msdataSM](#)” on page 15.)

If you have multiple samples you wish to compare, remember to set up a subdirectory for each sample. Ensure that each subdirectory contains all data files for a single sample.

You may add up to ten directory levels between **msdataSM** and your sample files, but shorter path lengths reduce memory usage.

- 3 If you have *.mgf files or appended *.pkl files, where each contains multiple spectra, copy or move them to your new subdirectory or subdirectories. IM *.pkl files are concatenated *.pkl files that contain all the spectra, so place them in the main folder rather than **cpick_in**. Proceed to [step 5](#). If you have individual *.pkl or *.dta spectral files, proceed to [step 4](#).
- 4 *Under* your new subdirectory, create another subdirectory called **cpick_in** and copy or move your files there. Note that you *must* use this subdirectory name!
- 5 After you transfer data files to the server, remove any spaces or parentheses in the data file names. Delete them or replace them with other characters, such as underscores.
- 6 Make sure you have both read and write permissions for each subdirectory that contains your sample files. (To check, right-click the folder and select **Properties**.) Make sure all user groups have full permissions. (Check the **Security** tab.)

To transfer MS-only files

NOTE

For Agilent TOF .d files and Q-TOF data in MS-only mode, you do *not* follow the workflow described below for other instrument files. You first generate a mass list with the Molecular Feature Extraction algorithm within the MassHunter Workstation Qualitative Analysis software, then copy the mass list from the compound table into the Spectrum Mill Manual PMF page, described in the *Application Guide*. See also Chapter 5 of the *Familiarization Guide* for exercises to learn how to search MS-only data.

- 1 Use your instrument data system to export calibrated, centroided, de-isotoped data as peak list files. Next you will transfer the peak lists to the Spectrum Mill server.
- 2 In the directory **msdataSM**, create a new subdirectory for your sample files. Name it anything you want, but do not use spaces or parentheses in the name. (See “[To access the server directory msdataSM](#)” on page 15).

Transferring Spectral Files to Spectrum Mill Server

To transfer MS-only files

You may add up to ten directory levels between **msdataSM** and your sample files, but shorter path lengths reduce memory usage and speed processing of large data sets.

- 3 *Under* your new subdirectory, create another subdirectory called **fit_batch_in** and copy or move your files there. Note that you *must* use this subdirectory name!
- 4 After you transfer data files to the server, remove any spaces or parentheses in the data file names. Delete them or replace them with other characters, such as underscores.
- 5 Make sure you have both read and write permissions for each subdirectory that contains your sample files. (To check, right-click the folder and select **Properties**.) Make sure all user groups have full permissions. (Check the **Security** tab.)

Starting the Spectrum Mill Software

Start the software from the server

- 1 Double-click the desktop icon to launch the Spectrum Mill workbench.
The icon is created during installation.
- 2 Ensure that you see the Spectrum Mill home page (Figure 2).



Start the software from the PC client

- 1 Double-click the desktop icon to launch the Spectrum Mill workbench.
 - a If you do not see this icon, start your Internet Explorer web browser on the client PC.
 - b In your browser window, type the address of the Spectrum Mill home page.
This has the general format **server/millhome.htm**. (If you don't know the URL, ask the person who installed the software or your system administrator.)
- 2 Ensure that you see the Spectrum Mill home page (Figure 2).



NOTE

If the Spectrum Mill Home shortcut is not on a client and you want to create it, do this:

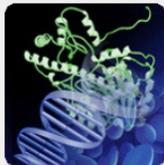
- a Right-click the title bar on the home page, and select **Create shortcut**.
- b Rename the Desktop icon to **Spectrum Mill Home**.

Or follow the instructions in the Tips and Techniques section of the Online Help entitled "To Create a Shortcut to the Spectrum Mill Workbench on Your Desktop".

Starting the Spectrum Mill Software

Start the software from the PC client

Agilent Technologies **Spectrum Mill MS Proteomics Workbench** Rev B.06.00



Help:

- Spectrum Mill Basics
- MS/MS Search
- Workflow Automation
- PMF Search
- Protein Databases
- Server Administration
- Tips and Tricks
- FAQs
- What's New

Useful Tables:

- Mutation Mass Shifts
- Dipeptide Masses

Manuals:

- Quick Start Guide
- Application Guide
- Familiarization Guide

Process Automation Tools	
Workflows	Automate data extraction, search, validation, and summary.
Completion Log	View log of completed extractions, searches, validations, and summaries
Request Queue	View status of request queue for submitted requests, monitor results output
Mass Spectral Interpretation Tools	
Data Extractor	Prepare MS/MS data files for Spectrum Mill processing
MS/MS Search	Search database with MS/MS spectra
Autovalidation	Autovalidate MS/MS spectra, calculate false-positive rates
Spectrum Matcher	Match MS/MS spectra against other MS/MS spectra
<i>de novo</i> Sequencing	Perform Sherenga <i>de novo</i> MS/MS spectral interpretation
Manual PMF Search	Search database with MS spectra from Peptide Mass Fingerprint data
Result Summary Tools	
Protein/Peptide Summary	Summarize results from MS/MS searches
Quality Metrics & FDR	Summarize metrics for data quality and FDR
Spectrum Summary	Summarize characteristics of MS/MS spectra
<i>de novo</i> Summary	Summarize results of Sherenga <i>de novo</i> MS/MS spectral interpretations
PMF Summary	Summarize results from PMF searches
Utilities	
Tool Belt	Collection of Spectrum Mill utility tools
Protein Databases	Manipulate FASTA sequence databases for use with Spectrum Mill software
Archive Data	Zip and Unzip dataset directories
Peptide Selector	Select peptides from protein digest likely to produce high quality MS/MS spectra
MRM Selector	Build/Export MRM transition lists based on experimental data in Spectrum Mill
Multiple Sequence Aligner	Run Clustal W and highlight non-identical AA's across the alignment
MS Digest	Predict peptide masses for enzymatic digestion of protein
MS Edman	Search database with text or partial peptide sequence
MS Product	Predict product ion masses from peptide sequence
MS Comp	Calculate AA compositions fitting parent mass and immonium ions
MS Isotope	Display isotope patterns of peptides
Peptide String Match	Search database with a list of peptide sequences
Peptide List to Masses	Convert a list of peptides to masses and formulas

Figure 2 Spectrum Mill home page

NOTE

To access help from the home page, click a topic on the left side of the page.

Using Normal and Iterative Workflows to Process Data

In this section you learn about the normal and iterative workflows to process multiple MS/MS data files with the Spectrum Mill MS Proteomics Workbench. Both workflows use tasks set up in Spectrum Mill modules (Figure 2). For an overview of the purpose of each module and where more information can be found for each one, see “Spectrum Mill modules and their purposes” on page 28.

Figure 3 shows two roadmaps for processing MS/MS data. For some samples and studies, it is sufficient to process only through a normal workflow, which reliably identifies most proteins in a sample with one round of processing. Detailed steps are covered in Chapter 1 of the *Application Guide*.

For others, you take advantage of an iterative workflow, which iterates through multiple rounds of database search and results validation, with the goal of gathering as much information as possible to better characterize the identified proteins.

The Spectrum Mill workbench provides a means to segregate search results that contain a valid interpretation of an MS/MS spectrum from those that do not. Results that are not validated can then be subjected to subsequent rounds of searches (against other databases or with a mass gap search, for example). Results that are validated can be summarized in a results table. These alternative steps are covered in Chapter 1 of the *Application Guide*.

Both workflows are summarized in “MS/MS workflow summary” on page 23. While the summary steps described here are representative of what many users want to do, there are certainly other paths through the software. The manual set and the online help give you the details you need to customize the workflows.

To use Spectrum Mill workbench for an MS-only workflow with a PMF search, see Chapter 6 of the *Application Guide*.

CAUTION

The default settings for the Spectrum Mill workbench have been carefully optimized. In general, you should change only the settings that are highlighted in red text. If you change other settings, be sure you are confident in your choices. If you wish to restore default settings, see the online help for the Spectrum Mill page of interest or load a parameter file from the “Defaults” parameter folder.

Using Normal and Iterative Workflows to Process Data

Start the software from the PC client

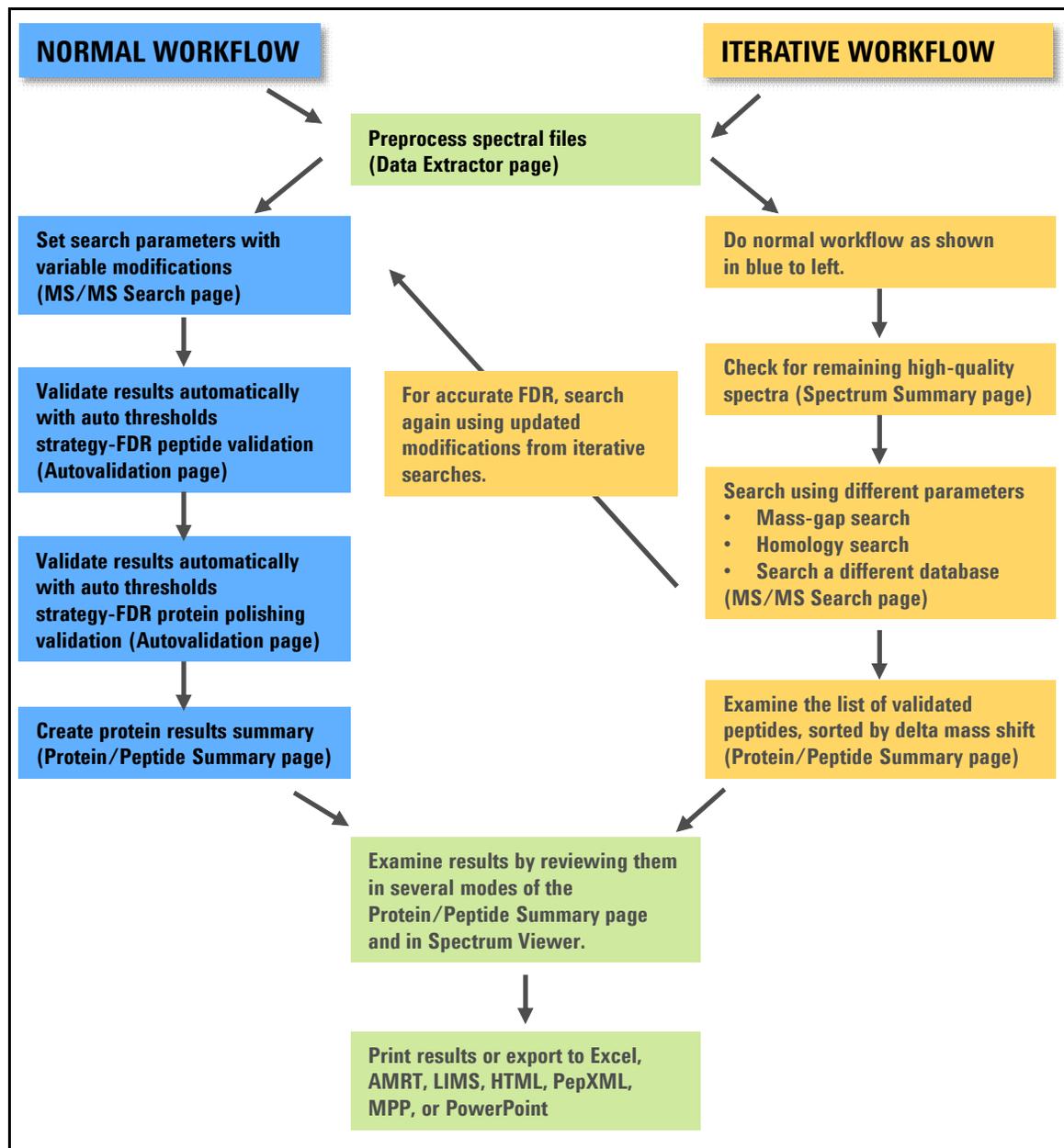


Figure 3 FDR-based workflows to process MS/MS data. FDR = false discovery rate.

MS/MS workflow summary

Figure 3 and the summary below describe the FDR-based normal and iterative workflows recommended for use with the Spectrum Mill workbench.

While the following instructions summarize the overall workflows, the step-by-step procedures are given in Chapter 1 of the *Application Guide*.

Before you begin processing, make sure you have set up a directory structure on the server that makes it easy to summarize and compare your results, and you have copied or moved the raw MS/MS data files to the Spectrum Mill server. See “[Transferring Spectral Files to Spectrum Mill Server](#)” on page 14.

Normal workflow

The normal workflow assumes that the goal is to reliably identify most proteins in the sample in only one round. To do this, you search first with variable modifications you think are present in your sample. For more details on choosing variable modifications, see Chapter 1 of the *Application Guide*.

- 1 Extract the data (Data Extractor page).
 - Set the correct cysteine modification.
 - For **Merge nearby MSⁿ scans with same precursor m/z**, if necessary change the time range to be compatible with your chromatographic data. For high resolution MS1 data, the precursor *m/z* value is ignored, and is based on the resolution indicated in the data file.
- 2 Search a database in **Variable Modifications** mode. Add modifications you think are present in your sample (MS/MS Search page).
- 3 Autovalidate the matches at a specified FDR, first in **Auto thresholds** strategy, **Peptide** mode, then (optionally) in **Auto thresholds** strategy, **Protein Polishing** mode. Use default settings or those in Chapter 1 of the *Familiarization Guide* (Autovalidation page).
 - Validation means that you accept the proposed database match as the correct match for the MS/MS spectrum.
- 4 Summarize and view the results (Protein/Peptide Summary page).

NOTE

If you are unsure of which modifications to search, a mass gap search can help. This search shows mass shifts of the modified peptides from the expected unmodified peptides, to help you identify modifications. Use a mass gap search in an iterative workflow after a variable modifications search and autovalidation. See “[Iterative workflow](#)” on page 24.

Once you identify the modifications you can use them as variable modifications in a normal workflow.

See Chapter 3 of the *Familiarization Guide* for exercises to help you do mass gap searches.

Iterative workflow

If your study requires you to characterize the identified proteins more extensively, use an iterative workflow. The steps you choose depend upon your sample, the available information in the databases you search, and how important it is to extract all available information from a sample set.

In a typical iterative workflow, you do a variable modifications search (as in the normal workflow). Then you search using different parameters. For example, you could do a homology search, a mass gap search (to find unexpected modifications), or search a different database. When you do a homology search or mass gap search, mark the check box to **Search previous hits**.

See Chapter 1 of the *Application Guide* and Chapters 2 and 3 of the *Familiarization Guide* for details about the possible steps you can use in an iterative workflow.

NOTE

To get an accurate false discover rate (FDR), all modifications must be searched in a single search. So if you do a search and later determine that other modifications are necessary, search again with *all* the modifications in a single search. Be sure to mark the check box to **Remove all prior MS/MS Search results**. Then autovalidate again.

Running Workflow Automation or Interactive Automation

This section gives introductory background for setting up and executing workflow automation and for running interactive automation.

Setting up and executing workflow automation

Workflow automation lets you place a series of tasks into a queue; then the Spectrum Mill workbench executes them in the proper order. For example, you can automate a typical workflow for MS/MS data files from protein digests:

- Data Extractor
- MS/MS Search
- Autovalidation
- Protein/Peptide Summary
- False discovery rate (FDR) and quality metrics
- Data archive

Workflow automation uses a Workflows page, an Edit Workflow page, and a Request Queue/Completion Log viewer. You access these pages through the Process Automation Tools section of the Spectrum Mill home page. See Chapter 3 of the *Application Guide* for instructions on how to set up and run automated workflows.

Set up workflow automation with parameter files

To create a workflow within the Edit Workflow page, you select task parameter files in the order you want them to execute during the workflow. A *parameter file* specifies the settings for each task. The first step to set up a workflow is to save the settings you want to use in parameter files.

Most Spectrum Mill pages let you save and load parameter files associated with the page. For example, Data Extractor, MS/MS Search (and PMF Search for MS data), Autovalidation, and Protein/Peptide Summary (and PMF Summary for MS data) include buttons to save settings in parameter files, and to load settings from parameter files. See Chapters 1 and 6 of the *Application Guide* for instructions on how to set up parameter files.

Running Workflow Automation or Interactive Automation

Setting up and executing workflow automation

For nearly all the pages, you use the parameter files, whose names are called tasks in the Workflows page, to build and execute workflows. Peptide Selector, MRM Selector, and Sherenga *de novo* Summary are exceptions. For these programs, it is convenient to save parameter files, but you do not use them in an automated workflow.

The Workflows and Edit Workflow pages let you view a task's page in read-only mode, with the parameters shown for the parameter file you select.

Execute multiple data folders in parallel

Workflow automation lets the Spectrum Mill workbench execute an ordered list of tasks for one or more data folders. All data folders must contain the same type of data (for example, all .raw, all .d, etc.). When Spectrum Mill workbench executes a workflow, it adds the tasks to a Request Queue, and executes them in order for a particular data folder.

After you select multiple data file folders and start the workflow, the Spectrum Mill workbench first does the Spectral Extraction and MS/MS Search on each of them independently, or if you have a multi-core CPU, it processes them in parallel. Then the program does the Autovalidation and Protein/Peptide Summary independently, or on the combination of the directories if you grouped proteins across multiple directories in the Autovalidation page.

Use the Request Queue/Completion Log Viewer

The Request Queue Viewer shows a list of all tasks that are currently executing and those that are queued for execution. It lists the tasks in the order they were queued. Because some tasks depend upon earlier tasks, the tasks that are currently executing do not always appear at the top of the list.

The Completion Log Viewer shows a list of all tasks that have completed, with most recent shown at the top. This log includes all queued requests, whether you queued them interactively or via a workflow. Two of the columns let you view additional information.

Running interactive automation

In the *Application Guide* see Chapter 1 for instructions on processing MS/MS data interactively, and Chapter 6 for MS data.

Interactively place tasks in the queue

The request queue concept extends to execution from a particular task's page. For example, in the Data Extractor page, the Extract button submits the extraction to the Request Queue, as does the Search button in the MS/MS and PMF Search pages.

Pages where interactive execution is helpful – Autovalidation and Protein/Peptide Summary for MS/MS data and PMF Summary for MS-only data – have a Queue request check box that is cleared by default. If you do not mark the check box, you see the results in the current window, as in earlier Spectrum Mill versions. If you mark the Queue request check box, the program puts the request into the queue, and saves the results output to an html file that you can view later.

Running interactive automation

You can execute portions of an MS/MS or MS workflow one step at a time, by executing requests from each task's page. For example, you may:

- 1 Use the Data Extractor page to select one or more data folders to be processed with a parameter file, and submit the request.
- 2 Go to the MS/MS Search page, where the current data folders are already selected, load a parameter file, then start the search.
- 3 Go to the Autovalidation page, select a parameter file and click the Queue request check box to queue the validation request for one or both autovalidation modes.
- 4 Go to Protein/Peptide Summary and queue one or more summary requests.

Quick Reference

Spectrum Mill modules and their purposes

The table below gives the purpose for each of the Spectrum Mill modules available on the home page.

Table 1 Spectrum Mill modules and their purposes

Tool type	Module	Function	Application Guide Chapter(s)
Workflow automation	Workflows	Create, edit and execute workflows, which are sequences of tasks (parameter files) for extractions, searches, validations and summaries.	3
	Completion Log	View log of completed extractions, searches, validations and summaries.	3
	Request Queue	View status of request queue for submitted requests; monitor results output.	3
Spectral preprocessing	Data Extractor	Extract spectra from data files. Processing depends on data type and may include averaging, centroiding, filtering by quality, assigning precursor charge, and calculating spectral features.	1, 6
Interpretation	MS/MS Search	Search MS/MS spectra against databases.	1, 6
	Autovalidation	Validate search hits based on various threshold strategies and specified false discovery rate cutoffs	1, 2
	Spectrum Matcher	Match MS/MS spectra against other MS/MS spectra	1
	<i>de novo</i> Sequencing	Use Sherenga <i>de novo</i> sequencing to interpret MS/MS spectra.	1, 4
	Manual PMF Search	Search MS-only spectra against databases.	6

Table 1 Spectrum Mill modules and their purposes (continued)

Tool type	Module	Function	Application Guide Chapter(s)
Result summary	Protein/Peptide Summary	Summarize, review, filter, and validate results from MS/MS searches.	1, 2
	Quality Metrics & FDR	Provide false discovery rate (FDR) and metrics for data quality.	1, 2
	Spectrum Summary	Sort and classify MS/MS spectra based on spectral characteristics.	1
	<i>de novo</i> Summary	Summarize and review <i>de novo</i> interpretations.	1, 4
	PMF Summary	Summarize and review MS-only results.	6
Utilities	Tool Belt	Use tools to speed processing, examine parameters and search statistics, stop processes, etc.	7
	Protein Databases	Create indices for databases so searches run faster, create subset databases, etc.	9
	Archive Data	Archive instrument-created files, search result files, spectral files and data directories	8
	Peptide Selector	Run theoretical digestion and then select peptides that meet certain criteria.	8
	MRM Selector	Build/Export MRM transition lists based on experimental data in Spectrum Mill.	8
	Multiple Sequence Aligner	Align amino acid sequences of proteins from a database and highlight the amino acids that differ	8
	MS Digest	Run theoretical enzymatic digestion and calculate masses of peptides that result.	8
	MS Edman	Search text fields in protein databases, with or without peptide mass filtering.	8
	MS Product	Calculate theoretical fragment ion masses from peptides.	8
	MS Comp	Calculate amino acid composition for a peptide, given peptide mass and immonium ions.	8
MS Isotope	Calculate and view isotope patterns of peptides.	8	

Quick Reference

Spectrum Mill modules and their purposes

Table 1 Spectrum Mill modules and their purposes (continued)

Tool type	Module	Function	Application Guide Chapter(s)
	Peptide String Match	Search a database with a list of peptide sequences	8
	Protein List to Masses	Calculates masses and formulas for a set of specified peptides	8

Interpreting scores from MS/MS Search

The following definitions and guidelines will help you evaluate your MS/MS search results. For detailed information, on your Spectrum Mill server, in the help for MS/MS Search, see the section on “Ranking/Scoring of Results.”

Score definitions

- Peptide score** Score of an individual peptide. This score reflects the information content (amount of useful fragmentation) in the MS/MS spectrum.
- Protein score** Score of the overall protein, calculated by adding scores of all the peptides detected for the protein
- Scores greater than 15 for Agilent Q-TOF and ion trap instruments almost always represent valid results.
- % SPI** Percentage of the extracted MS/MS ion current explained by theoretical fragmentation of the database hit

Guidelines for validating high-resolution data based on MS/MS Search scores

Because the mass accuracy for instruments like the Agilent Q-TOF is better than that of the ion trap, you search with a narrower mass tolerance. Therefore, there is a better chance that lower-scoring results are valid.

Table 2 Guidelines for validation of high-resolution results

Peptide score	Quality	Peptide fragmentation	Likelihood of valid interpretation
> 15	Outstanding, well-assigned	Thorough	When combined with % SPI of 60 or greater, very likely to be valid
> 9	Good	Substantial	When combined with % SPI of 60 or greater, likely to be valid
> 5	Mixed quality	Moderate	Review results to determine whether interpretation is valid; more likely to be valid than same scores for ion trap data
< 5	Generally poor	Less	Not likely to be valid

When you manually validate, apply the following filters in Protein/Peptide Summary:

Quick Reference

Interpreting scores from MS/MS Search

- Peptide score > 5
- % SPI > 60

Guidelines for validating low-resolution data (such as ion trap) based on MS/MS Search scores

Table 3 Guidelines for validation of low-resolution results

Peptide score	Quality	Peptide fragmentation	Likelihood of valid interpretation
> 15	Outstanding, well-assigned	Thorough	When combined with % SPI of 70 or greater, very likely to be valid
> 10	Good	Substantial	When combined with % SPI of 70 or greater, likely to be valid
> 5	Mediocre	Few ions, look for mostly b's or y's	Review results to determine whether interpretation is valid; less likely to be valid than same scores for Q-TOF data
< 5	Poor	Little to none	Not likely to be valid

Protein/Peptide Summary Modes

Table 4 Protein/Peptide Summary Modes

If you want to:	And you want results organized by:	Then use this mode:	Example application
Validate results	Peptides	Peptide-Spectrum Match	Manual review and validation of MS/MS search results, organized by peptide
	Proteins, then peptides	Protein Summary Details	Manual review and validation of MS/MS search results, organized by protein
Summarize results by proteins	Proteins, then samples	Protein-Protein Comparison	Compare two or more samples, each of which may contain multiple fractions.
	Proteins, then peptides	Protein Summary Details	View proteins, with supporting peptide details
Summarize results by peptides	Peptides only	Peptide-Spectrum Match	List of all peptides identified in the data
	Peptides, then samples	Protein-Peptide Distribution Columns	Method development (evaluation of 2D LC/MS/MS or other fractionation scheme)
	Rows: peptides clustered by protein Columns: samples	Protein-Peptide Comparison	Table of distinct peptides, case-insensitive (CI). Collapse all peptide spectrum matches (PSMs) that contain same peptide into one row. Combine quantitation for each peptide: median of all PSMs.
	Distinct peptides only	Peptide-Distinct	Provides filtering options to view peptides with variable modifications. Collapse peptide list to best-scoring representative PSM per distinct peptide.
Summarize results by variable modification sites	Rows: variable modification sites clustered by protein Columns: samples	Protein- Var Mod Site Comparison	Compare variable modifications across samples. Provides table of phosphorylation sites. Collapse all PSMs that contain same phosphorylation site into one row. Combine quantitation for each site: median of all PSMs.
View a list of proteins identified via a single peptide	Peptides only	Protein-Single Peptide ID	Examination of results where a single peptide was used to generate a protein identification

www.agilent.com

In this Book

The *Quick Start Guide* presents first steps to use the Spectrum Mill MS Proteomics Workbench.

© Agilent Technologies, Inc. 2016

Revision A, October 2016



G2721-90055

Revision A0



Agilent Technologies